

CENTRO REGIONAL UNIVERSITARIO CÓRDOBA IUA



“BUSINESS INTELLIGENCE CON HEFESTO” ING. EN INFORMÁTICA – BASES DE DATOS II PROYECTO

PROFESOR

Ing. Caffaratti, Melisa Anabella

INTEGRANTES

Carrizo, Matías

Guillaumet, Gabriel Agustín

Llamosas, Simon

10 de junio de 2024



Índice General

Índice General.....	1
1. Introducción.....	2
2. Descripción de la problemática.....	3
3. Metodología Hefesto.....	4
3.1 Análisis de Requerimientos.....	4
3.1.1 Preguntas de Negocio.....	4
3.1.2 Indicadores y Perspectivas.....	5
3.2 Análisis de Data Sources.....	6
3.2.1 Hechos e Indicadores.....	6
3.2.2 Mapeo.....	6
3.2.3 Granularidad.....	8
3.2.4 Modelo Conceptual Ampliado.....	9
3.3 Modelo Lógico del DataWarehouse.....	9
3.3.1 Tipología.....	9
3.3.2 Tablas de Dimensiones.....	9
3.3.3 Tabla de Hechos.....	11
3.3.4 Uniones.....	12
3.4 Integración de Datos.....	12
3.4.1 Carga Inicial.....	12
Descripción del proceso.....	12
Procedimientos.....	14
3.4.2 Actualización.....	16
4. Diagrama Entidad - Relación.....	17
5. Diccionario de datos.....	18
6. Reportes y dashboards.....	22
Estadística de Temporadas.....	23
Análisis de Liga por Tiempo.....	24
Estadística de Liga.....	25
Análisis de Enfrentamientos.....	26
Clásico.....	27
Estadística y Análisis de Equipos.....	28
7. Bibliografía y recursos.....	30



1. Introducción

En el contexto actual, donde la toma de decisiones basadas en datos es fundamental para el éxito empresarial, la implementación de soluciones de Inteligencia de Negocios (BI) se ha convertido en una necesidad imperativa. Este trabajo presenta una solución de Inteligencia de Negocios utilizando la metodología Hefesto, una guía comprensiva y estructurada para el desarrollo de sistemas de BI efectivos, que se aplicará sobre un conjunto de datos basados en partidos de ligas de fútbol europeas.

2. Descripción de la problemática

El fútbol, como deporte global, genera una cantidad masiva de datos que, si se analizan correctamente, pueden ofrecer insights valiosos para equipos, jugadores, analistas y aficionados. Con el creciente interés en el análisis de datos y la aplicación de técnicas exhaustivas en el deporte, se presenta una oportunidad única para explotar estos datos en la toma de decisiones estratégicas y operativas.

A pesar de la disponibilidad de una gran cantidad de datos relacionados con el fútbol, existe una brecha significativa en cómo estos datos se utilizan para obtener ventajas competitivas. Los equipos de fútbol, los analistas deportivos y otros interesados enfrentan desafíos para:

1. **Análisis de enfrentamiento:** Obtener datos históricos en los distintos partidos disputados en las ligas europeas, permitiendo analizar los encuentros desde distintas perspectivas.
2. **Optimización de Estrategias de Juego:** Identificar patrones y tendencias en los datos que puedan influir en la estrategia de juego, alineaciones y tácticas en los distintos enfrentamientos.
3. **Predicción de Resultados de Partidos:** Mejorar la precisión en la predicción de resultados ante un partido, así tomar decisiones a la hora de plantear el modo de juego.

Implementar un análisis de datos robusto, utilizando la metodología Hefesto, permitirá abordar estas problemáticas de manera efectiva. La metodología Hefesto proporciona un marco estructurado para el desarrollo de sistemas de Inteligencia de Negocios, garantizando una documentación clara y comprensible que facilita la replicabilidad y la mejora continua.

Algunos beneficios que guían el propósito del análisis de datos en el fútbol:

- Mejora de la Toma de Decisiones: Equipos y entrenadores pueden basar sus decisiones en datos concretos, mejorando la planificación de partidos y las tácticas.
- Desarrollo de Estrategias: Los equipos contarán con la información necesaria para establecer estrategias personalizadas de cara a un enfrentamiento.



3. Metodología Hefesto

3.1 Análisis de Requerimientos

3.1.1 Preguntas de Negocio

Dado que nuestro análisis no es realizado sobre una empresa particular, sino que está enfocado a la investigación deportiva, hemos hablado con algunos contactos informados sobre el análisis futbolístico para informarnos sobre qué datos recabar para concretar una evaluación que tenga valor para el negocio. A partir de esto, hemos determinado las siguientes preguntas a responder, relacionadas al análisis de los distintos encuentros europeos, para la construcción de nuestro datawarehouse :

- ¿Cuántos goles se convierten en cada enfrentamiento?
- ¿Cuál es la tendencia de goles totales anotados en cada temporada y liga?
- ¿Cuál es la frecuencia de tarjetas rojas en cada temporada y liga?
- ¿Qué enfrentamientos son los más aguerridos como locales y como visitantes en cada liga?
- ¿Qué equipo en el enfrentamiento ganó más puntos como local y como visitante en cada temporada y liga?
- ¿Cuáles son los clásicos europeos más “peleados”? ¿Quiénes tienen a favor la estadística?
- ¿Qué equipo tiene más eficiencia de cara al arco, en un enfrentamiento?
- ¿Qué equipo tiene el mayor promedio de goles por partido en casa y como visitante en cada liga?
- ¿Qué equipos son los más goleadores a nivel histórico? ¿Y cuáles son aquellos que más goles reciben?
- ¿Qué equipos reciben más amonestaciones en cada liga?

Siguiendo la metodología Hefesto, podríamos reformular las preguntas para verlas desde una perspectiva más enfocada al negocio, analizando el rendimiento de cada equipo en los distintos enfrentamientos, jugando tanto de local como de visitante. De esto, obtenemos lo siguiente:

- Cantidad de goles por equipo en cada temporada y liga.
- Cantidad de tarjetas por equipo en cada temporada y liga.
- Cantidad de clásicos ganados por equipo en cada liga a nivel histórico.
- Cantidad de goles por tiro de cada equipo, por cada liga y temporada.
- Cantidad de puntos por equipo en cada temporada y liga.



Mientras que las cuestiones no resueltas serán analizadas en reportes y dashboards en la sección 6 del presente informe, dada la complejidad de los datos a analizar y la necesidad de una visión más amplia y sencilla de los mismos.

3.1.2 Indicadores y Perspectivas

A partir de las preguntas planteadas con respecto al negocio, identificamos los siguientes indicadores y sus respectivas perspectivas de análisis:

Indicadores - Perspectivas

- Cantidad de goles totales por encuentro entre equipo local y equipo visitante en cada temporada y liga.
- Cantidad de tarjetas amarillas por encuentro entre equipo local y equipo visitante en cada temporada y liga.
- Cantidad de tarjetas rojas por encuentro entre equipo local y equipo visitante en cada temporada y liga.
- Promedio de tarjetas rojas en cada temporada y liga.
- Promedio de tarjetas amarillas en cada temporada y liga.
- Promedio de goles por temporada y liga.
- Cantidad de puntos ganados por encuentro entre equipo local y equipo visitante en cada temporada y liga.

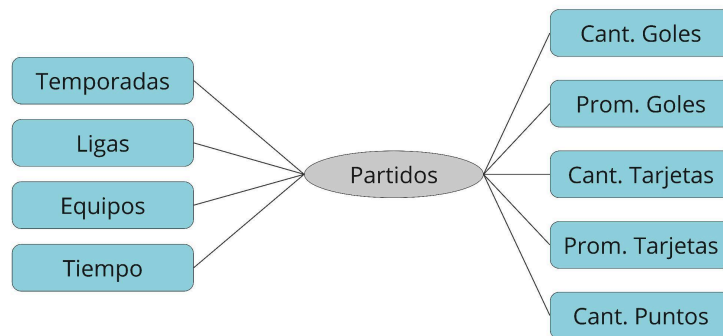
Que resumiendo, nos queda en lo siguiente

- Indicadores
 - Cantidad de goles por partido
 - Cantidad de tarjetas amarillas por partido
 - Cantidad de tarjetas rojas por partido
 - Cantidad de puntos ganados
- Perspectivas
 - Temporadas
 - Ligas
 - Equipos
 - Tiempo

Siendo estos los elegidos dada la importancia estratégica que tienen dichos factores en la previa de un encuentro futbolístico.

3.1.3 Modelo Conceptual

A partir de las perspectivas e indicadores planteados en la sección anterior, construimos el precedente modelo conceptual:



3.2 Análisis de Data Sources

3.2.1 Hechos e Indicadores

A continuación, describiremos la forma de cálculo para cada uno de los indicadores propuestos en la sección anterior:

Indicador: cantidad de goles por enfrentamiento

- **Hechos:** goles totales
- **Función de agregación:** SUM

Indicador: cantidad de tarjetas por enfrentamiento (diferenciado entre amarillas y rojas)

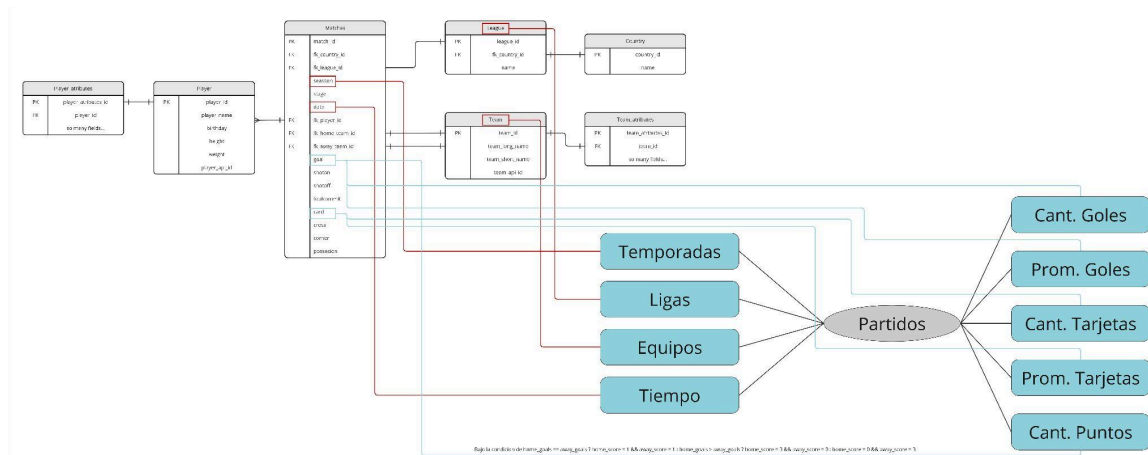
- **Hechos:** amarillas totales
- **Función de agregación:** SUM
- **Hechos:** rojas totales
- **Función de agregación:** SUM

Indicador: cantidad de puntos (dependiendo si se quieren de local o de visitante)

- **Hechos:** puntos local / visitante
- **Función de agregación:** SUM

3.2.2 Mapeo

A partir del diagrama entidad - relación original de la base de datos OLTP, el [mapeo](#) realizado fue el siguiente:



Donde:

- La dimensión “Temporadas” se obtiene de la columna “season” de la tabla “Matches”
- La dimensión “Ligas” se obtiene de la tabla “League”
- La dimensión “Equipos” se obtiene de la tabla “Teams”
- La dimensión “Tiempo” se obtiene de la columna “date” de la tabla “Matches”
- El indicador “Cant. goles” se obtiene a partir de la columna “goals” de la tabla matches (con un respectivo tratamiento de datos que se explicará más adelante) sumando los goles del enfrentamiento. Su fórmula de cálculo será:

$$\text{SUM}(\text{goles_local} + \text{goles_visitante})$$

- El indicador “Prom. Goles” se obtiene a partir de la columna “goals” de la tabla matches (con un respectivo tratamiento de datos que se explicará más adelante) haciendo el promedio de los goles por enfrentamiento. Su fórmula de cálculo será:

$$\text{AVG}(\text{goles_local} + \text{goles_visitante})$$

- El indicador “Cant. tarjetas” se obtiene a partir de la columna “cards” de la tabla matches, sumando las tarjetas del enfrentamiento. Su fórmula de cálculo será:
 - Para las tarjetas amarillas:

$$\text{SUM}(\text{amarillas_local} + \text{amarillas_visitante})$$

- Para las tarjetas rojas:

$$\text{SUM}(\text{rojas_local} + \text{rojas_visitante})$$

- El indicador “Prom. tarjetas” se obtiene a partir de la columna “cards” de la tabla matches, haciendo el promedio de las tarjetas del enfrentamiento. Su fórmula de cálculo será:



- Para las tarjetas amarillas.

$AVG(amarillas_local + amarillas_visitante)$

- Para las tarjetas rojas.

$AVG(rojas_local + rojas_visitante)$

- El indicador “Cant. puntos” se obtiene a partir de dos columnas agregadas a la tabla “Matches” donde se indican los puntos adjudicados por el equipo local en una y los puntos adjudicados por el visitante en la otra. Su fórmula de cálculo será:

$SUM(puntos_local)$ para los puntos obtenidos de local

$SUM(puntos_visitante)$ para los puntos obtenidos de visitante.

3.2.3 Granularidad

En base a lo anterior, y analizando la información relevante para el negocio sobre cada perspectiva, obtenemos las siguientes tablas:

Team

name: nombre completo del equipo.

Season

name: nombre de la temporada en el formato Inicio/Fin.

Inicio: representa el año de inicio de la temporada.

Fin: representa el año de fin de la temporada.

League

name: nombre completo de la liga.

Time

year: año de la fecha.

semester: semestre del año.

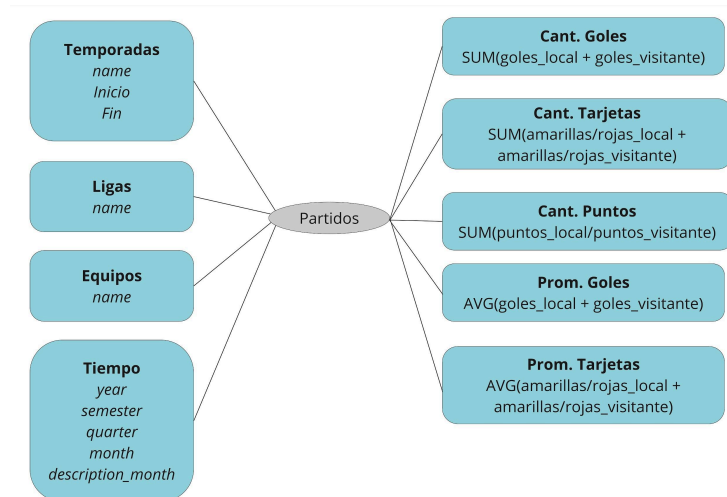
quarter: cuatrimestre del año

month: mes del año.

description_month: el mes pero representado con su nombre.

3.2.4 Modelo Conceptual Ampliado

En base a las secciones anteriores, obtenemos el siguiente [modelo conceptual ampliado](#):



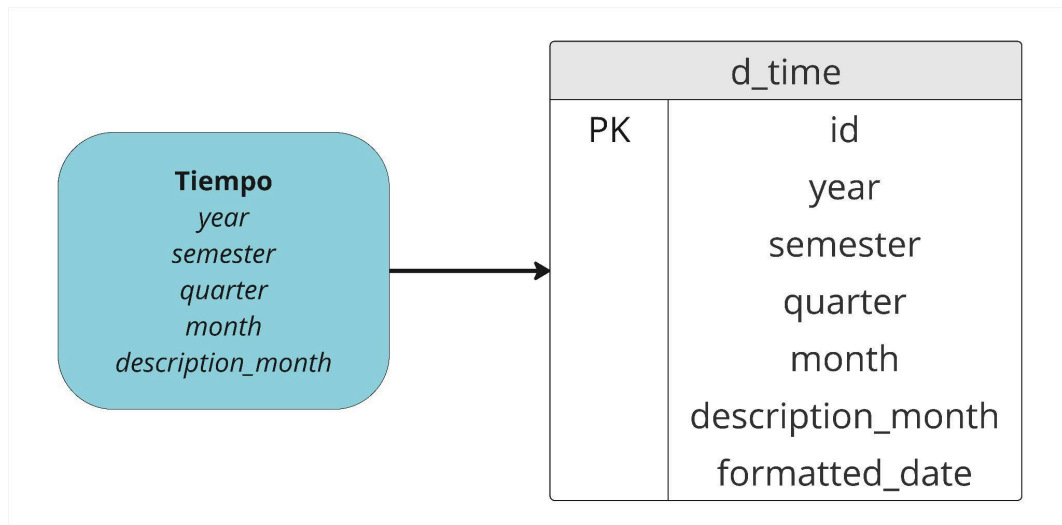
3.3 Modelo Lógico del Data Warehouse

3.3.1 Tipología

Para la construcción de nuestro data warehouse, optamos por un esquema de estrella, dado su óptimo funcionamiento para el análisis multidimensional y la baja complejidad de las relaciones entre nuestras perspectivas.

3.3.2 Tablas de Dimensiones

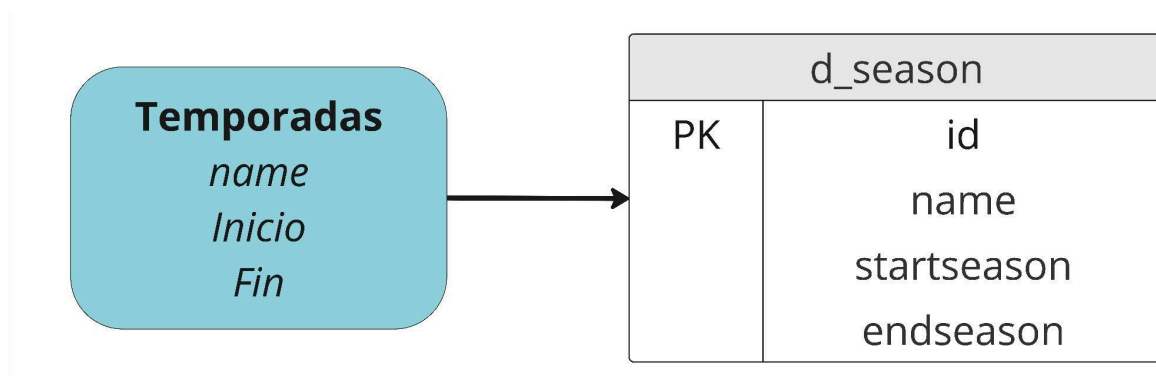
A partir de las perspectivas planteadas, se armaron las siguientes tablas de dimensiones



Donde se agregaron los campos:

id: clave primaria de la dimensión. Representa unívocamente a una fecha.

formatted_date: fecha en formato AAAA-MM-DD HH-MM-SS

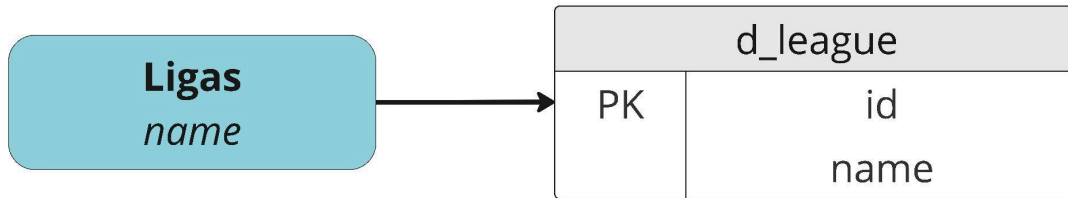


Donde se agregaron los campos:

id: clave primaria de la dimensión. Representa unívocamente a una temporada.

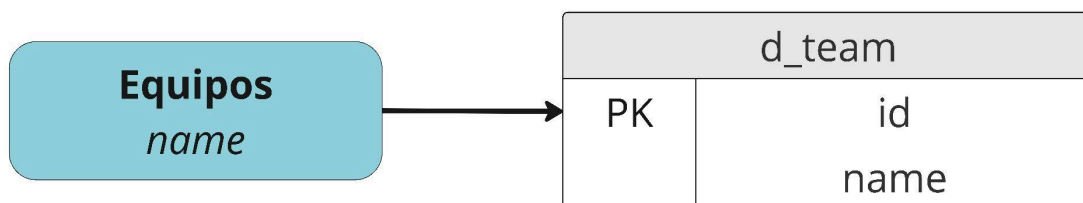
startseason: se renombró el campo “inicio” a “startseason”

endseason: se renombró el campo “fin” a “endseason”



Donde se agregaron los campos:

id: clave primaria de la dimensión. Representa unívocamente a una liga.



Donde se agregaron los campos:

id: clave primaria de la dimensión. Representa unívocamente a un equipo.

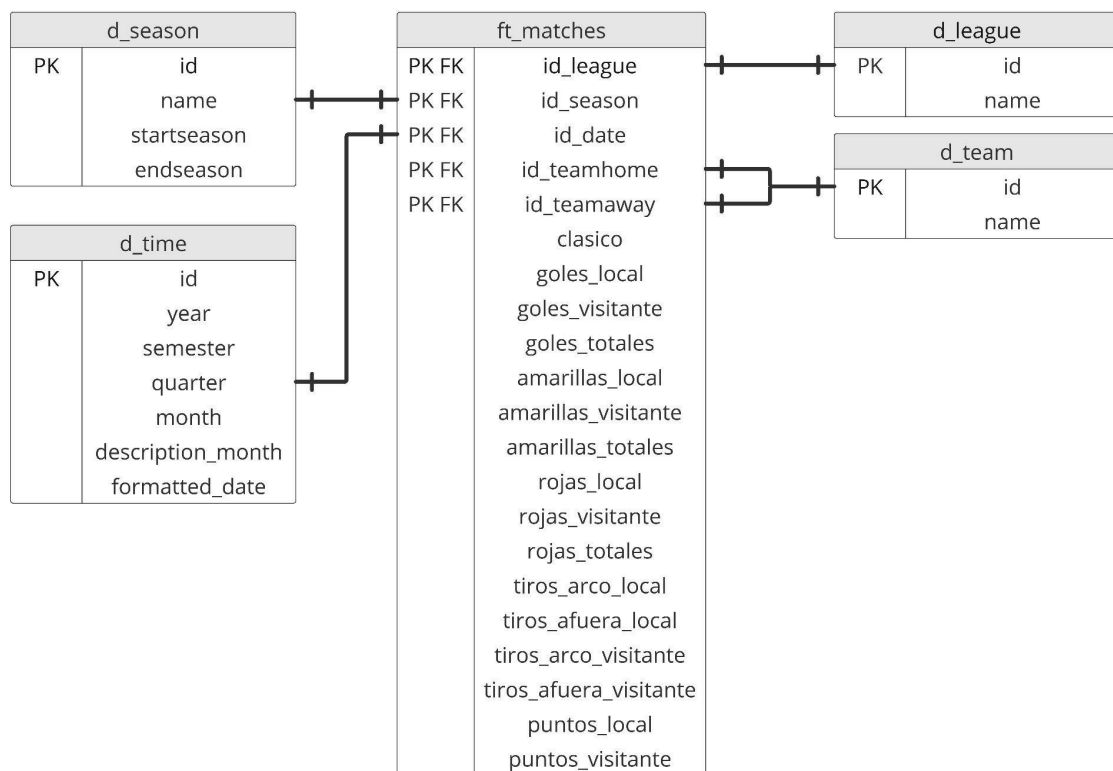
3.3.3 Tabla de Hechos

A raíz de las dimensiones planteadas en la sección anterior, y los hechos previamente definidos, se establece la siguiente [tabla de hechos](#):

ft_matches	
PK FK	id_league
PK FK	id_season
PK FK	id_date
PK FK	id_teamhome
PK FK	id_teamaway
	clasico
	goles_local
	goles_visitante
	goles_totales
	amarillas_local
	amarillas_visitante
	amarillas_totales
	rojas_local
	rojas_visitante
	rojas_totales
	tiros_arco_local
	tiros_afuera_local
	tiros_arco_visitante
	tiros_afuera_visitante
	puntos_local
	puntos_visitante

3.3.4 Uniones

A continuación, mostramos cómo se relacionan dimensiones con la tabla de hechos mediante las claves foráneas definidas:



3.4 Integración de Datos

3.4.1 Carga Inicial

Descripción del proceso

Como fuente de datos, utilizamos un dataset consistente en una base de datos SQLite correspondiente a datos de partidos de 11 ligas europeas desde la temporada 2008/09 a la 2015/16, el cual fue obtenido del siguiente enlace: [data source](#).

Analizando los datos provistos por el dataset, este originalmente contaba con tablas correspondientes a ligas, equipos, jugadores y atributos varios, junto a una tabla con información de los partidos, la cual incluye relaciones a las anteriores, además de agregar datos sobre el partido como la temporada en la que se jugó, la fecha y, además, los goles,



tarjetas, faltas cometidas, corners e información extra que no es relevante para nuestro análisis; donde los mismos estaban provistos en columnas con formato XML

En base a estas columnas de formato XML, realizamos tablas temporales con la información que contenían:

Tabla de Temporadas (NO TEMPORAL)

En este caso, el dato no era un XML, sino un campo string, pero se optó por hacer una tabla aparte dado que la consideramos como una perspectiva de análisis para el negocio.

- id
- nombre
- año de inicio
- año de fin

Tabla de goles

- id del partido
- id del equipo que convirtió

Tabla de amonestaciones

- id del partido
- id del equipo que recibió la amonestación
- tipo de amonestación

Tabla de tiros al arco

- id del partido
- id del equipo

Tabla de tiros afuera

- id del partido
- id del equipo

Tabla de faltas

- id del partido
- id del equipo

Donde cada equipo tendrá la cantidad de filas correspondientes a la cantidad de instancias de los distintos hechos correspondientes a un partido, insertas en estas tablas.



Procedimientos

Para construir estas tablas temporales, utilizamos scripts de Python, implementados con Pandas, donde dichos scripts levantan una conexión a la base de datos SQLite, ejecutan una query SQL para seleccionar las columnas con los datos XML y, posteriormente, lanzan otra sobre la misma conexión para insertar los datos obtenidos en la nueva tabla temporal.

Una vez confeccionadas las tablas temporales, ya podemos comenzar a armar el esquema (*schema*) 'tmp'. Este proceso se realizó utilizando Pentaho, mediante su solución *data integration*, la cual nos permite migrar los datos desde el data source (OLTP) hasta el datawarehouse (tmp).

Perspectivas

Mediante el uso de Pentaho Data Integration, realizamos la conexión al OLTP, donde limpiaremos los datos que no son relevantes o de utilidad para el proceso de toma de decisiones del negocio.

Por otro lado, creamos una nueva conexión para la base de datos PostgreSQL que contiene el esquema tmp. Es a partir de esta que comenzamos el proceso de extracción.

Por cada una de las tablas en nuestra OLTP, generamos una transformación que nos permitirá migrar los datos desde la misma hacia nuestro esquema. Esto se realizó mediante las prestaciones de Data Integration.

Proceso del pasaje de datos

Mediante la herramienta de “entrada de tabla”, hacemos una query a la base de datos OLTP, extrayendo los datos que consideramos necesarios para, posteriormente, y mediante la herramienta “salida de tabla”, dirigir el flujo de datos hacia las nuevas tablas del esquema tmp. Este proceso se repitió para cada una de las perspectivas.

Para el proceso de desarrollo de la tabla de hechos, lo que se hizo fue tomar la información de las nuevas tablas creadas en el esquema tmp y mapearla con la tabla correspondiente a la tabla de hechos de la OLTP, así como también, el cálculo de las distintas métricas:

- clásico
- goles del equipo local
- goles del equipo visitante
- goles totales del encuentro (suma de las dos anteriores)
- amarillas del equipo local
- amarillas del equipo visitante
- amarillas totales del encuentro (suma de las dos anteriores)
- rojas del equipo local
- rojas del equipo visitante



- rojas totales del encuentro (suma de las dos anteriores)
- tiros al arco del equipo local
- tiros al arco del equipo visitante
- tiros fuera del equipo local
- tiros fuera del equipo visitante

Una vez realizado esto, completamos el esquema tmp para pasar al esquema dsa.

En este nuevo esquema buscamos establecer un puente entre la tmp y el dw. Esto se realiza mediante la adición de claves subrogadas para cada una de las dimensiones. El objetivo de dichas claves es independizarse de la OLTP y sanitizar los datos.

Partiendo de esta idea, el armado de las dimensiones consistió en traer la misma información colocada en el esquema tmp, pero agregando esta nueva clave subrogada, independiente de los identificadores ya presentes en la tabla.

Una vez armadas las dimensiones, el procedimiento para la tabla de hechos es muy similar que en la tmp, con la diferencia de que se utilizan los id originales del OLTP para mapear las perspectivas con la tabla de hechos, pero como foreign key en la tabla de hechos se coloca la clave subrogada.

Una particularidad, es el agregado de la dimensión tiempo, la cual fue construida mediante una transformación, a partir del dato existente en la OLTP, en el formato “AAAA - MM - DD HH:MM:SS”.

Mediante la ejecución de dicha transformación en Data Integration mas un script, se tomó dicho formato, y por medio del tratamiento de datos, conformamos la tabla con los siguientes campos:

Dimensión Tiempo:

- iddate (AAAA:MM:DD)
- year
- semester
- quarter
- month
- description month
- formatted date (en el formato original)

A partir de esta tabla, se realizó el mapeo de la dimensión tiempo con la tabla de hechos, por medio de la formatted date, colocando en esta última, el iddate (correspondiente en nuestro caso a la clave subrogada de la dimensión tiempo) como foreign key.

Una vez realizado esto, completamos el esquema dsa para pasar al esquema dw.



Para esto, construimos las dimensiones a partir de las existentes en la dsa, con la particularidad de que, en este punto, eliminamos las claves originales (provenientes de la OLTP) y mantenemos únicamente las subrogadas.

Hecho esto, mapeamos la tabla de hechos con las dimensiones definidas en el esquema dw de igual manera que se definió en el esquema anterior.

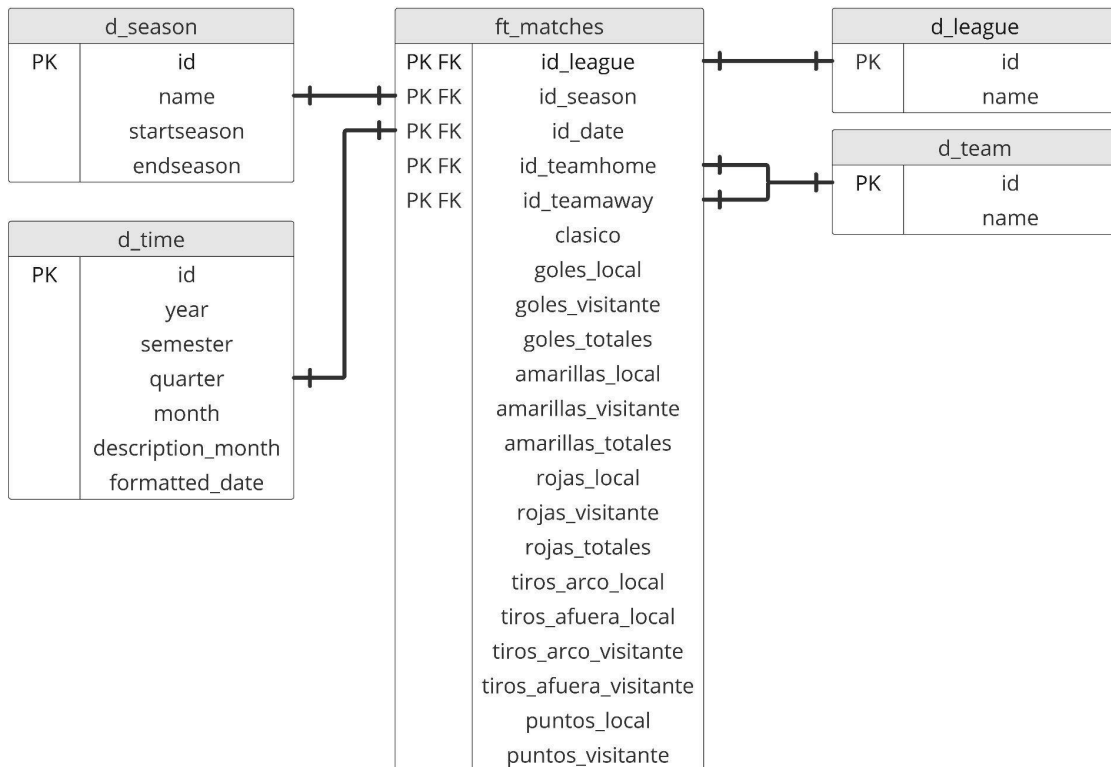
Con este último paso, concretamos la carga inicial de datos a nuestro data warehouse.

3.4.2 Actualización

Si bien en este proceso no se realizan actualizaciones, bastaría con definir las variables *date_from* y *date_to* para las migraciones de datos, estableciendo *date_from* como la fecha de concretado de la carga inicial, y como *date_to* la fecha hasta el punto en que se quieran cargar datos. Posterior a esto, ejecutar el proceso ETL completo insertaría los nuevos datos en nuestro data warehouse.

4. Diagrama Entidad - Relación

Si bien el mismo ya fue expuesto en la metodología Hefesto, es de suma importancia observar cómo se unen las dimensiones con la tabla de hechos:



Podemos observar que:

- Clave primaria de la tabla de hechos: la clave primaria está compuesta por las relaciones a las dimensiones, es decir, la primary key (PK) de la tabla de hechos se compone de las foreign key (FK) de las dimensiones.
- Claves foráneas: cada dimensión cuenta con una clave foránea en la tabla de hechos.

5. Diccionario de datos

Tabla “d_leagues”					
Columna	Descripción	Tipo de Dato	Rango de Datos	Fórmula	Observaciones
id	representa unívocamente a la liga	Integer	-	-	Primary Key NOT NULL
name	nombre de la liga	Varchar(100)	-	-	NOT NULL

Tabla “d_seasons”					
Columna	Descripción	Tipo de Dato	Rango de Datos	Fórmula	Observaciones
id	representa unívocamente a la temporada	Integer	-	-	Primary Key NOT NULL
name	nombre de la temporada	Varchar (9)	-	-	NOT NULL
startseason	año de inicio de la temporada	Integer	-	-	NOT NULL
endseason	año de fin de la temporada	Integer	-	-	NOT NULL

Tabla “d_teams”					
Columna	Descripción	Tipo de Dato	Rango de Datos	Fórmula	Observaciones
id	representa unívocamente al equipo	Integer	-	-	Primary Key NOT NULL
name	nombre del equipo	Varchar(50)	-	-	NOT NULL

Tabla “d_time”					
Columna	Descripción	Tipo de Dato	Rango de Datos	Fórmula	Observaciones
id	representa unívocamente a la fecha	Integer	-	-	Primary Key NOT NULL
year	año de la fecha	Integer	-	-	NOT NULL
semester	semestre del año	Varchar (20)	-	-	NOT NULL
quarter	cuatrimestre del año	Varchar (20)	-	-	NOT NULL
month	mes del año	Integer	-	-	NOT NULL
description_month	nombre del mes	Varchar (10)	-	-	NOT NULL

Tabla “ft_matches”					
Columna	Descripción	Tipo de Dato	Rango de Datos	Fórmula	Observación
idleague	representa unívocamente la liga del partido	Integer	-	-	Primary Key Foreign Key NOT NULL
idseason	representa unívocamente la temporada del partido	Integer	-	-	Primary Key Foreign Key NOT NULL
iddate	representa unívocamente la fecha del partido	Integer	-	-	Primary Key Foreign Key NOT NULL
idteamhome	representa unívocamente al equipo local del partido	Integer	-	-	Primary Key Foreign Key NOT NULL
idawayhome	representa unívocamente al equipo visitante del partido	Integer	-	-	Primary Key Foreign Key NOT NULL
clasico	año de la fecha	Integer	-	-	NOT NULL
goles_local	goles del equipo local	Integer	-	-	NOT NULL
goles_visitante	goles del equipo visitante	Integer	-	-	NOT NULL
goles_totales	goles del encuentro	Integer	-	goles_local + goles_visitante	NOT NULL



amarillas_local	amarillas del equipo local	Integer	-	-	NOT NULL
rojas_local	rojas del equipo local	Integer	-	-	NOT NULL
amarillas_visitante	amarillas del equipo visitante	Integer	-	-	NOT NULL
rojas_visitante	amarillas del equipo local	Integer	-	-	NOT NULL
amarillas_total	amarillas del partido	Integer	-	amarillas_local + amarillas_visitante	NOT NULL
rojas_total	rojas del partido	Integer	-	rojas_local + rojas_visitante	NOT NULL
tiros_arco_local	tiros al arco del equipo local	Integer	-	-	NOT NULL
tiros_afuera_local	tiros afuera del equipo local	Integer	-	-	NOT NULL
tiros_arco_visitante	tiros al arco del equipo visitante	Integer	-	-	NOT NULL
tiros_afuera_visitante	tiros afuera del equipo visitante	Integer	-	-	NOT NULL
puntos_local	puntos obtenidos por el equipo local	Integer	-	-	NOT NULL
puntos_visitante	puntos obtenidos por el equipo visitante	Integer	-	-	NOT NULL

6. Reportes y dashboards

A través de esta sección, proporcionaremos un análisis profundo de varios aspectos clave del desempeño de los equipos en las ligas europeas, utilizando herramientas avanzadas de Business Intelligence como Power BI. Este análisis nos permitirá responder preguntas críticas sobre el rendimiento de los equipos en diferentes contextos, ofreciendo insights valiosos para clubes, analistas y aficionados.

Los reportes y dashboards que se presentan a continuación responden a preguntas específicas sobre tres áreas principales:

1. **Clásicos y Enfrentamientos:** Estadísticas históricas sobre los enfrentamientos entre equipos y las rivalidades más grandes de Europa. Aquí se podrán analizar los enfrentamientos más emblemáticos entre equipos, proporcionando datos detallados sobre victorias, empates y derrotas, goles convertidos, puntos obtenidos, etc. a lo largo de los años.
2. **Tendencias:** Análisis de las estadísticas en base a ligas en distintos períodos de tiempo. Exploraremos cómo fueron evolucionando los equipos a lo largo del tiempo en sus correspondientes ligas.
3. **Efectividad:** Cuáles son los equipos con más efectividad de cara a puerta. En esta parte del análisis, se mostrará qué equipos tienen las mejores tasas de conversión de tiros a gol, destacando la precisión y el rendimiento ofensivo de cada equipo en el enfrentamiento. Los datos incluirán métricas como tiros a gol, goles marcados y porcentaje de efectividad, ofreciendo una perspectiva clara sobre los equipos más letales frente al arco.
4. **Agresividad:** Qué equipos son los más rústicos dentro de la cancha. Esta sección se enfocará en las estadísticas relacionadas con el juego físico y las infracciones, mostrando qué equipos reciben más tarjetas y participan en más disputas físicas. Es una herramienta útil para identificar los equipos que emplean un estilo de juego más físico y agresivo.

Algunos ejemplos de reportes y dashboards que realizamos en estos análisis son los siguientes:

Estadística de Temporadas

En la siguiente imagen, vemos las estadísticas generales de un equipo para una liga y temporada específica.

Este reporte responde a las siguientes preguntas:

- ¿Qué equipos son los más goleadores a nivel histórico? ¿Y cuales son aquellos que más goles reciben?
- ¿Qué equipos reciben más amonestaciones en cada liga?

Permitiéndonos un análisis detallado del rendimiento de los equipos para una temporada específica.

LIGAS					TEMPORADAS			
Belgium Jupiler League	England Premier League	France Ligue 1	Germany Bundesliga	Italy Serie A	2008/2009	2010/2011	2012/2013	2014/2015
					2009/2010	2011/2012	2013/2014	2015/2016

TABLA GENERAL DE LA LIGA PARA CADA TEMPORADA										
Equipo	Puntos	PJ	Ganados	Empatados	Perdidos	Goles Convertidos	Goles Recibidos	DG	Amarillas	Rojas
FC Barcelona	94	38	30	4	4	120	30	90	66	1
Atlético Madrid	90	38	29	6	3	65	19	46	90	1
Real Madrid CF	89	38	28	5	5	112	40	72	71	2
Athletic Club de Bilbao	67	38	20	11	7	62	46	16	85	4
RC Celta de Vigo	63	38	18	11	9	54	61	-7	113	3
Villarreal CF	60	38	17	12	9	43	37	6	99	1
Málaga CF	53	38	15	15	8	44	37	7	108	1
Sevilla FC	55	38	15	13	10	54	51	3	101	2
Real Sociedad	45	38	12	17	9	46	50	-4	106	2
SD Eibar	44	38	12	18	8	53	62	-9	106	1
Rayo Vallecano	42	38	11	18	9	56	74	-18	112	6
RCD Espanyol	41	38	11	19	8	42	75	-33	127	0
Real Betis Balompié	44	38	11	16	11	37	54	-17	110	0

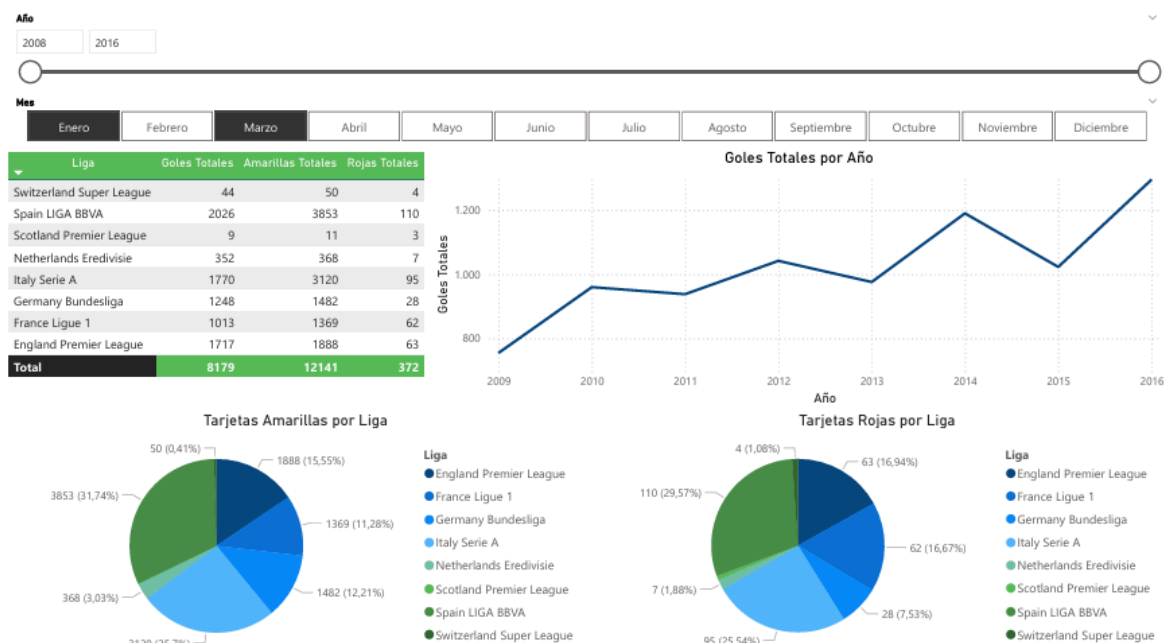
Análisis de Liga por Tiempo

La siguiente imagen muestra un análisis de la cantidad de goles convertidos por un equipo en una liga en específico, junto con la cantidad de tarjetas amarillas y rojas que obtuvo en un tiempo específico.

Este dashboard nos ayuda a responder preguntas como:

- ¿Cuál es la tendencia de goles totales anotados en cada temporada y liga?
- ¿Cuál es la frecuencia de tarjetas en cada temporada y liga?

Pudiendo realizar un análisis más a fondo de las ligas en distintos periodos de tiempo.

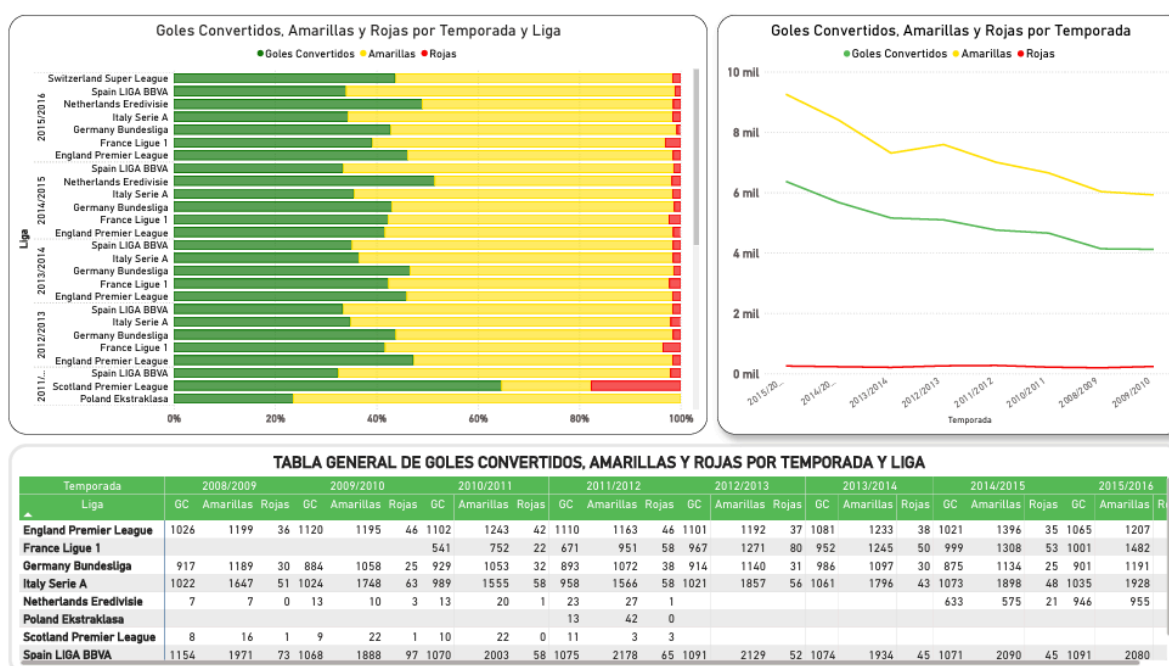


Estadística de Liga

Este reporte responde a preguntas como:

- ¿Cuál es la tendencia de goles totales anotados en cada temporada y liga?
- ¿Cuál es la frecuencia de tarjetas en cada temporada y liga?

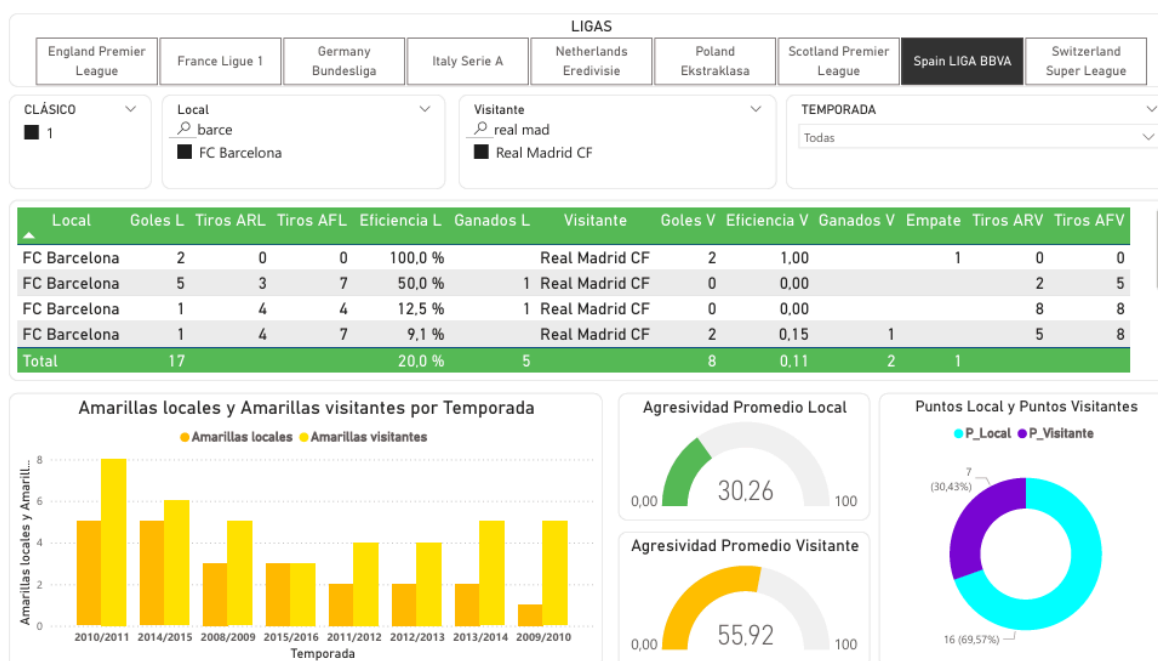
Pudiendo realizar un análisis comparativo de las grandes ligas europeas.



Análisis de Enfrentamientos

La siguiente imagen muestra un análisis de los enfrentamientos entre dos equipos. Este reporte responde a las siguientes preguntas:

- ¿Cuántos goles se convierten en cada enfrentamiento?
- ¿Qué equipo en el enfrentamiento ganó más puntos como local y como visitante en cada temporada y liga?
- ¿Qué enfrentamientos son los más aguerridos como locales y como visitantes en cada liga?



Clásico

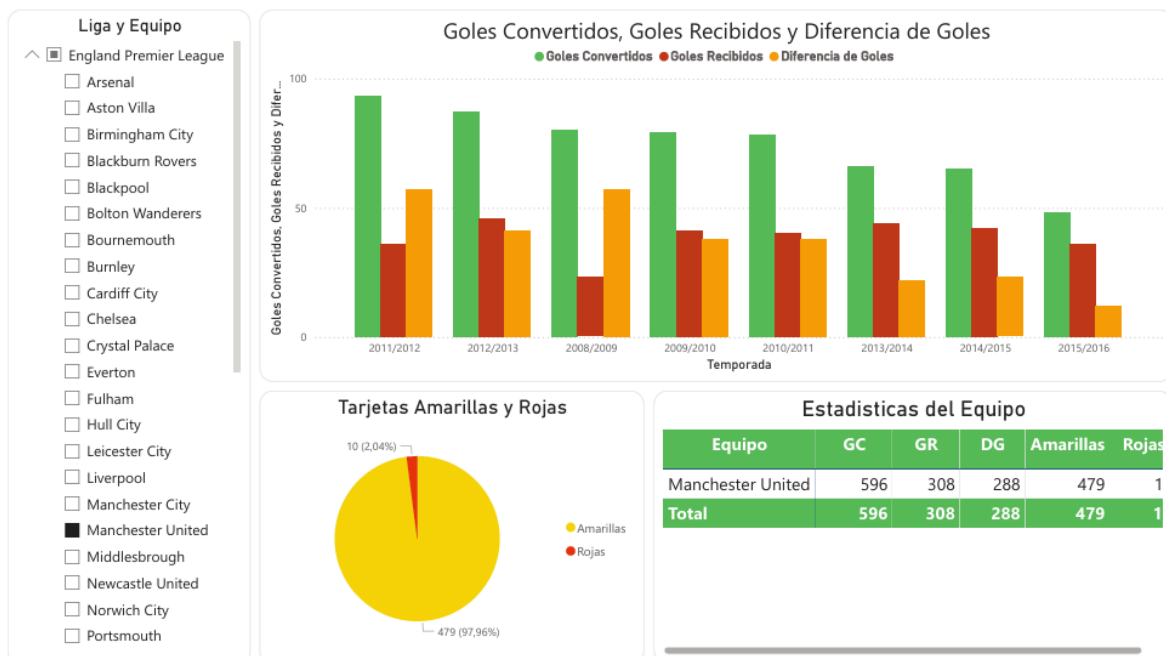
El siguiente reporte muestra estadísticas de los enfrentamientos “clásicos” europeos. Esto nos ayuda a responder a la pregunta de “¿Cuáles son los clásicos europeos más peleados? y ¿Quiénes tienen a favor la estadística?”

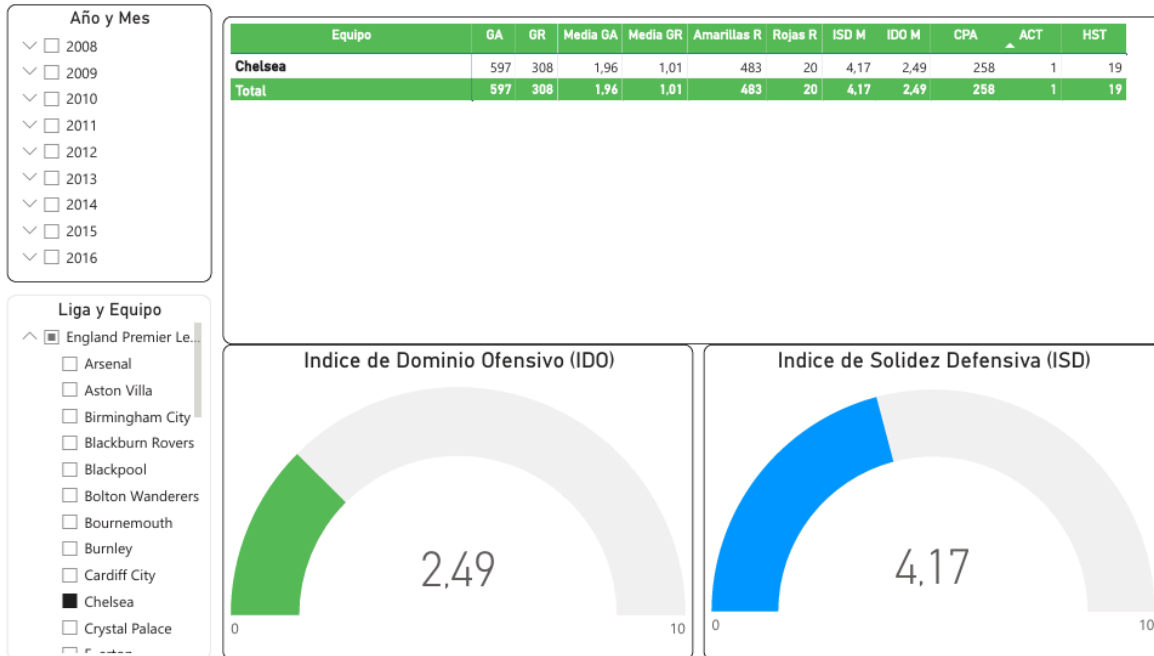


Estadística y Análisis de Equipos

En los siguientes reportes, nos centraremos específicamente en equipos concretos, evaluando su eficiencia en el campo de juego. El siguiente reporte permite responder a preguntas como:

- ¿Qué equipo tiene más eficiencia de cara al arco, en un enfrentamiento?
- ¿Qué equipo tiene el mayor promedio de goles por partido en casa y como visitante en cada liga?
- ¿Qué equipos son los más goleadores a nivel histórico? ¿Y cuales son aquellos que más goles reciben?
- ¿Qué equipos reciben más amonestaciones en cada liga?





Se adjunta el link desde donde pueden verse todos los reportes y dashboard de forma más clara: [Dashboards y Reportes](#)



7. Bibliografía y recursos

Bernabeu, D., & García, M. (2017). Hefesto Data Warehousing: guía completa de aplicación teórico-práctica; metodología Data Warehouse.

Link del repositorio en GitHub: <https://github.com/simonl14/dw-db2>.